| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 09/715.552 | 11/17/2000 | Bhaskar Ghosh | 50277-0406 | 6519 |

| | | |
|---|---|---|
| 29989 7590 12/14/2004 | | EXAMINER |
| HICKMAN PALERMO TRUONG & BECKER, LLP | | ZHEN, LI B |
| 1600 WILLOW STREET | | |
| SAN JOSE, CA 95125 | ART UNIT | PAPER NUMBER |
| | 2126 | |

DATE MAILED: 12/14/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

|  | Application No. | Applicant(s) |
|---|---|---|
| **Office Action Summary** | 09/715,552 | GHOSH ET AL. |
|  | Examiner | Art Unit | |
|  | Li B. Zhen | 2126 | |

*-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --*

## Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) FROM
THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed
  after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any
  earned patent term adjustment. See 37 CFR 1.704(b).

## Status

1)☒ Responsive to communication(s) filed on <u>22 July 2004</u>.

2a)☐ This action is **FINAL**.          2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is
closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

## Disposition of Claims

4)☒ Claim(s) *1-26* is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☐ Claim(s) _____ is/are rejected.

7)☒ Claim(s) *8-12 and 21-25* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

## Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

## Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All   b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage
application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

## Attachment(s)

1)☒ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
    Paper No(s)/Mail Date _____.

4)☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____ .

5)☐ Notice of Informal Patent Application (PTO-152)

6)☐ Other: _____.

## DETAILED ACTION

1.      Claims 1 – 26 are pending in the current application.

### *Allowable Subject Matter*

2.      Claims 21 – 25 are objected to as being dependent upon a rejected base claim,

but would be allowable if rewritten in independent form including all of the limitations of

the base claim and any intervening claims.

3.      Claims 8 – 12 would be allowable if rewritten to overcome the rejection(s) under

35 U.S.C. 101, set forth in this Office action and to include all of the limitations of the

base claim and any intervening claims.

### *Response to Arguments*

4.      Applicant's arguments with respect to the claims have been considered but are

moot in view of the new ground(s) of rejection.

### *Claim Rejections - 35 USC § 101*

5.      35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of
> matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the
> conditions and requirements of this title.

6.      Claims 1 – 13 are rejected under 35 U.S.C. 101 because they are directed to

non-statutory subject matter.

7.      Claims 1 – 13 are directed to method steps which can be practiced mentally in

conjunction with pen and paper, therefore they are directed to non-statutory subject

matter.  Specifically, as claimed, it is uncertain what performs each of the claimed

method steps.  Moreover, each of the claimed steps, inter alia, randomly selecting,

determining, assigning and distributing, can be practiced mentally in conjunctions with

pen and paper.  The claimed steps do not define a machine or computer implemented

process [see MPEP 2106].  Therefore, the claimed invention is directed to non-statutory

subject matter.  (The examiner suggests applicant to change "method" to "computer

implemented method" in the preamble to overcome the outstanding 35 U.S.C. 101

rejection).

## Claim Rejections - 35 USC § 102

8.      The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that

form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

9.      **Claims 1, 13, 14 and 26 are rejected under 35 U.S.C. 102(b) as being**

**anticipated by Gal et al., USPN 5,729,732 (hereinafter Gal, cited in the previous**

**office action).**

10.     As to claim 1, Gal teaches a method for distributing data items from a particular

set of data into a plurality of buckets based on distribution keys associated with said

data items (Gal, see Abstract, "A method is described for operating a computer to sort a

set of data records each having an associated key for governing the sort process"), the

method comprising the steps of:

randomly selecting data items from said particular set of data to produce a

sampled set of data items (Gal, col.3, lines 51-52, "The random sampling can be

achieved, for example, by taking a predetermined set of n indices");

determining a range for each bucket of the plurality of buckets based on a

number of the distribution keys associated with the sampled set of data items that fall

within the range (Gal, col. 5, lines 40 – 52 and col. 6, lines 7 – 15; more than half the

keys in a sample to be sorted share the same 4-byte prefix (300), the data is split into

three sets (310), one for the keys with the common prefix, one for the keys with a prefix

which is smaller than the common prefix, and one for the keys with a prefix which is

larger than the common prefix);

assigning said plurality of ranges to said plurality of buckets (Gal, see Abstract,

"defining a plurality of buckets, each bucket corresponding to a respective one of a

plurality M of subintervals in the range"); and

distributing each data item in said particular set of data to the bucket that has

been assigned the range into which falls the distribution key of the data item (Gal, see

Abstract, "distributing the keys among the buckets by determining directly from each key

value the index of the subinterval into which the key value falls").

11.    As to claim 13, Gal teaches determining ranges that contain an approximately

equal amount of distribution keys associated with said sampled set of data items (Gal,

col.2. lines 7-19, "the index of the subinterval into which each key falls is determined

directly from the key value. This means that the distribution of each key into the

respective bucket can be performed in a time, which does not depend on the number of

buckets used in the distribution. The subintervals are equal").

12.    As to claims 14 and 26, these are product claims that correspond to method

claims 1 and 13; note the rejection to claims 1 and 13 above, which also meet these

product claims.

### Claim Rejections - 35 USC § 103

13.    The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set
> forth in section 102 of this title, if the differences between the subject matter sought to be patented and
> the prior art are such that the subject matter as a whole would have been obvious at the time the
> invention was made to a person having ordinary skill in the art to which said subject matter pertains.
> Patentability shall not be negatived by the manner in which the invention was made.

14.    **Claims 2 – 4, 7, 15 – 17 and 20 are rejected under 35 U.S.C. 103(a) as being**

**unpatentable over Gal in view of "Improved Histograms for Selectivity Estimation**

**of Range Predicates" (hereinafter Poosala).**

15.     As to claim 2, Gal teaches the invention substantially as claimed including

randomly selecting data items from a particular set of data to produce a sampled set of

data items (Gal, col. 3, lines 51 – 52);

determining a plurality of ranges based on the distribution keys associated with

the sampled set of data items (Gal, see Abstract, "determining a range for the key

values by sampling the key values");

assigning the plurality of ranges to the plurality of buckets (Gal, see Abstract,

"defining a plurality of buckets, each bucket corresponding to a respective one of a

plurality M of subintervals in the range"); and

distributing each data item in said particular set of data to the bucket that has

been assigned the range into which falls the distribution key of the data item (Gal, see

Abstract, "distributing the keys among the buckets by determining directly from each key

value the index of the subinterval into which the key value falls").

16.     Although Gal teaches the invention substantially, Gal does not specifically

teaches randomly selecting data items from each subset of a plurality of subsets of said

particular set of data to produce a sampled set of data items.

However, Poosala teaches selecting data items from each subset of a plurality of

subsets of said particular set of data to produce a sampled set of data items (Poosala,

i.e. Section 7.1 and 7.2, pp. 299 - 300; reservoir sampling, obtains a random sample of

size n during a single pass through the relation, p. 300, left col., lines 20 – 30).

17.     It would have been obvious to a person of ordinarily skilled in the art at the time

of the invention to apply the teaching of randomly selecting data items from each subset

of a plurality of subsets of said particular set of data to produce a sampled set of data

items as taught by Poosala to the invention of Gal because this does not require the

database system to support individual retrieval of randomly selected pages and hence

can be implemented in most current systems (p. 300, left col., 35 – 37).

18.    As to claim 3, Gal teaches randomly selecting data items from each partition of a

partitioned table (Gal, col.3, lines 51-54, "The random sampling can be achieved, for

example, by taking a predetermined set of n indices, for example from a pseudo-

random table, and picking the corresponding elements").

19.    As to claim 4, Gal teaches randomly selecting data items from subsets of data,

stored in buffers in volatile memory (i.e. R.AM), that represent results of one of more

previously performed operations (Gal, see F1G.1, col.3, lines 12-13, "the data

processing system which may be utilized for implementing the method and system of

the present invention includes a processor 10, a random access memory (RAM) 12, a

read only memory (ROM) 14, at least one non-volatile storage device 15, a computer

display monitor 16 and a keyboard 18").

20.    As to claim 7, Gal as modified teaches a method for distributing data items from

a particular set of data into a plurality of buckets based on distribution keys associated

with said data items (Gal, see Abstract, "A method is described for operating a

computer to sort a set of data records each having an associated key for governing the

sort process"), the method comprising the steps of:

randomly selecting data items from said particular set of data to produce a

sampled set of data items (Gal, col.3, lines 51-52, "The random sampling can be

achieved, for example, by taking a predetermined set of n indices");

wherein the particular set of data is durably stored on a plurality of durable

storage units (at least one non-volatile storage device 15; col. 3, lines 1 – 5 of Gal;

examiner notes that the term "durable storage unit" is very broad because any storage

medium are generally durable, therefore a "durable storage unit" would read on a

memory location of a storage device); and

the step of randomly selecting data items includes randomly selecting durable

storage units from said plurality of durable storage units and using the data items stored

on said randomly selected durable storage units as the sampled set of data items

(Poosala, i.e. Section 7.1 and 7.2, pp. 299 - 300; reservoir sampling, obtains a random

sample of size n during a single pass through the relation, p. 300, left col., lines 20 –

30);

determining a plurality of ranges based on the distribution keys associated with

the sampled set of data items (Gal, see Abstract, "determining a range for the key

values by sampling the key values");

assigning said plurality of ranges to said plurality of buckets (Gal, see Abstract,

"defining a plurality of buckets, each bucket corresponding to a respective one of a

plurality M of subintervals in the range"); and

distributing each data item in said particular set of data to the bucket that has

been assigned the range into which falls the distribution key of the data item (Gal, see

Abstract, "distributing the keys among the buckets by determining directly from each key

value the index of the subinterval into which the key value falls").

21.    As to claims 15 – 17 and 20, these are product claims that correspond to method

claims 2 – 4 and 7; note the rejection to claims 2 – 4 and 7 above, which also meet

these product claims.

22.    **Claims 5 and 18 are rejected under 35 U.S.C. 103(a) as being unpatentable**

**over Gal in view of Ogi (USPN 5,854,938, cited in previous office action).**

23.    As to claim 5, Gal teaches a method for evenly distributing data items to

corresponding buckets (Gal, see Abstract, CW method is described for operating a

computer to sort a set of data records each having an associated key for governing the

sort process, the method comprising determining a range for the key values by

sampling the key values; defining a plurality of buckets, each bucket corresponding to a

respective one of a plurality M of subintervals in the range").

However, Gal does not particularly teach processing the buckets with plural

processors concurrently operating in parallel to execute a task.

24.    Ogi teaches assigning the plurality of buckets to a plurality of processes (Ogi,

see F1G.3); and causing each process of said plurality of processes to perform, in

parallel with the other processes of said plurality of processes, an operation on the data items contained in any buckets assigned to the process (Ogi, see FIG.3).

Gal and Ogi are analogous art because they are from the same field of endeavor in parallel operations and data distribution. It would have been obvious for one of ordinary skill in the art at the time the invention was made to modify the method for sorting a set of data in a computer system using buckets having associated key values within a range of Gal such that it utilizes the parallel computer system having plural processors concurrently operating in parallel to execute tasks as teach d by Ogi.

25.     As to claim 18, this is a product claim that corresponds to method claim 5; note the rejection to claim 5 above, which also meet this product claim.

26.     **Claims 6 and 19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Gal and Poosala further in view of Marks (USPN 5,748,844, cited in previous office action).**

27.     As to claim 6, Gal teaches a method for randomly selecting data items from each subset of a plurality of subsets of said particular set of data (Gal, col.3, lines 40-41 and 46, "The file to be sorted comprises N records each of which has an associated key" and "A random sample of the keys is taken from the file $y.sub.1, y.sub.2, ...y.sub.n$").

However, Gal does not particularly teach selecting a distinct random seed for each subset of the plurality of subsets of said particular set of data.

28.     Marks teaches selecting a distinct random seed for each subset of the plurality of subsets of said particular set of data (Marks, F1G.3, element 32, col.3, lines 44-46, "the seed-growth heuristic initially assigns a small number of randomly chosen nodes to each part of the partition; these are the seed nodes").

Gal and Marks are analogous art because they are from the same field of endeavor in partitioning that can be applied to database design and parallel processing. Generally, random number generators have to be "seeded" with an initial seed. The use of random seed generators from which a "first" random number is derived is well known in the art.

It would have been obvious for one of ordinary skill in the art at the time the invention was made to modify the method for randomly selecting data items from each subset of a plurality of subsets such that it utilizes the random seed generator of Marks.

One of ordinary skill in the art would be motivated to do so because it assures the best random sampling of data for partitioning thereby yielding a superior sampling operation performed in parallel.


29.     As to claim 19, this is a product claim that correspond to method claim 6; note the rejection to claim 6 above, which also meet this product claim.


*Conclusion*

30.     Any inquiry concerning this communication or earlier communications from the examiner should be directed to Li B. Zhen whose telephone number is (571) 272-3768. The examiner can normally be reached on Mon - Fri, 8:30am - 5pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Meng-Ai An can be reached on (571) 272-3756. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Li B. Zhen
Examiner
Art Unit 2126

lbz

MENG-AI T. AN
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100